Computation of the Semantics of Autoepistemic Belief Theories [1]

Stefan Brass [a] Jürgen Dix [b] Teodor C. Przymusinski [c,2]

[a] *School of Information Sciences, University of Pittsburgh, PA 15260, USA*

[b] *Dept. of Computer Science, University of Maryland, A. V. Williams Build., College Park, MD 20752-3255, USA*

[c] *Dept. of Computer Science, University of California, Riverside, CA 92521, USA*

Abstract

Recently, one of the authors introduced a simple and yet powerful non-monotonic knowledge representation framework, called the *Autoepistemic Logic of Beliefs*, *AEB*. Theories in *AEB* are called *autoepistemic belief theories*. Every belief theory T has been shown to have the *least static expansion* $\overline{T}$ which is computed by iterating a natural monotonic *belief closure operator* Ψ_T starting from T. This way, the least static expansion $\overline{T}$ of any belief theory provides its natural non-monotonic semantics which is called the *static semantics*.

It is easy to see that if a belief theory T is *finite* then the construction of its least static expansion $\overline{T}$ stops after *countably* many iterations. However, a somewhat surprising result obtained in this paper shows that the least static expansion of any finite belief theory T is in fact obtained by means of a *single iteration* of the belief closure operator Ψ_T (although this requires T to be of a special form, we also show that T can be always put in this form). This result eliminates the need for multiple iterations in the computation of static semantics and allows us to replace the fixed-point definition of static semantics by the equivalent explicit and straightforward definition given by $\overline{T} = \Psi_T(T)$.

The second, closely related result establishes an intriguing relationship between static semantics $\overline{T}$ Clark's completions $comp(T)$ of finite belief theories. Here we use a slightly generalized version of $comp(T)$ (see Definition 3.2). It shows that the static semantics $\overline{T}$ of T is obtained by augmenting T with the set $\mathcal{B}comp(T) = \{\mathcal{B}F : F \in comp(T)\}$ thus ensuring that all formulae that belong to Clark's completion $comp(T)$ of T are believed to be true.

Both results open the way for more efficient implementation of static semantics: the first, because only one iteration is needed, and the second because reasoning in a non-standard logic (belief theories under static semantics) can be reduced to classical theorem proving involving Clark's completion.

Keywords: Non-Monotonic Reasoning, Logics of Knowledge and Beliefs, Semantics of Logic Programs and Deductive Databases.

1 Introduction

The problem of finding inference mechanisms, capable to model human commonsense reasoning and to reason in the absence of complete information, is one of the major research and implementation problems in AI. Accordingly, over the last 20–30 years a large number of formalizations of *commonsense* (or *non-monotonic*) reasoning, and their inference engines, have been proposed. Among these, the most prominent and most actively studied turned out to be Reiter's CWA (26), McCarthy's circumscription (18), Moore's autoepistemic logic (20), Reiter's default logic (25) and Kowalski's logic programming (17).

While logic programming is often just viewed as a formal specification for declarative programming languages, such as *PROLOG*, in the authors' view and in the view of many researchers actively studying this area, logic programming plays a much more important role as a *non-monotonic knowledge representation language*. Its importance stems from several factors. First of all, logic programming is a *relatively simple* non-monotonic formalism (compared to other major formalisms): it allows only atomic heads of clauses and implements only one non-monotonic feature, namely, default negation (or negation as failure) and it is restricted to atomic formulae. On the other hand, as shown by the extensive research conducted during the last decade or two, logic programming is *sufficiently expressive* to allow formalization of many important problems in commonsense reasoning. As a result, logic programming offers *fertile testing grounds* for studying, relating and comparing of various formalizations of commonsense reasoning in a relatively simple and yet quite expressive framework.

The results of such extensive studies have been quite impressive. A number of different semantics for logic programming have been proposed and studied in detail. Two of them turned out to be most promising: the *stable semantics* (15) and the *well founded semantics* (31). The relationship of the stable semantics to other major non-monotonic formalisms has been well established. It turns out that, under a suitable embedding, static semantics can be simply viewed as a special case of Moore's autoepistemic logic AEL (20).

Recently, these semantics have also been used in various areas of traditional

[1] Preliminary results of this note appeared in the Proceedings of the Fifth International Conference on Principles of Knowledge Representation and Reasoning (KR'96), L. C. Aiello, J. Doyle and S. C. Shapiro (editors), 1996, pp. 529–541.
[2] Partially supported by the National Science Foundation grant #IRI-9313061.

AI, like *planning* ((9)), *agent reasoning* ((28; 13; 14; 29; 10)) or in *verification* (see references in (30)). In particular it has been shown that systems based on these semantics (like smodels (21) or XSB (24)) can successfully compete with dedicated systems (the recent workshop on *Logic based AI* organized by Jack Minker, June 1999, gave impressive evidence for this).

Consequently, a further analaysis of how these semantics can be extended, unified and computed in a feasible way will most probably lead to new applications or better algorithms in the above mentioned or similar areas of AI. Our approach can be viewed as extending traditional semantics to semantics dealing with *beliefs*. Beliefs play an important role whenever available information is incomplete and can be described using statements expressing beliefs.

Our investigations show that the fixpoint of a certain sequence defining the static semantics (which itself contains most of the traditional nonmonotonic semantics) is (1) reached after *finitely* many steps, and (2) it is closely related to a version of Clarks completion for finite belief theories. Thus our results may also have an impact on those semantics.

We now explain our results in more detail. In (22; 23), one of the authors introduced a simple and yet powerful non-monotonic knowledge representation framework which isomorphically contains all of the above mentioned non-monotonic formalisms and semantics as special cases and yet is significantly more expressive than each one of these formalisms considered individually. The new logic, called the *Autoepistemic Logic of Beliefs*, AEB, has been shown to constitute a powerful new formalism which can serve as a *unifying framework* for several major non-monotonic formalisms. It allows us to better understand mutual relationships existing between different formalisms and semantics and enables us to provide them with simpler and more natural definitions. It also naturally leads to new, even more expressive, flexible and modular formalizations and semantics (23).

Theories in the Autoepistemic Logic of Beliefs are called *autoepistemic belief theories.* As it is the case in Moore's autoepistemic logic, AEL, semantics of belief theories is defined by means of introducing a class of their so called *static expansions.* Static expansions are fixed points of a natural monotonic *belief closure operator* Ψ_T. However, as opposed to autoepistemic logic and its stable expansions, every belief theory T has the *least static expansion* $\overline{T}$ which can be computed by iterating the belief operator Ψ_T. This way, the least static expansion of any belief theory provides its natural non-monotonic semantics which is called the *static semantics*, namely the set of all formulae contained in the least static expansion $\overline{T}$.

It is easy to see that if a belief theory T is *finite* then the construction of its least static expansion $\overline{T}$ will stop after *countably* many iterations. However, a powerful and somewhat surprising result obtained in this paper shows that the least static expansion of any finite belief theory T is in fact obtained by means of a *single iteration* of the belief closure operator Ψ_T.[3] We will later show that T can be always put in this form. This result eliminates the need for multiple iterations in the computation of static semantics and allows us to replace the fixed-point definition of static semantics by the equivalent explicit and straightforward definition given by $\overline{T} = \Psi_T(T)$. Needless to say, the existence of an equivalent *non-fixed point* definition of static semantics significantly simplifies this notion and the underlying theory. It also provides the foundation for the interesting results and applications obtained in (5).

The second, closely related result establishes a very interesting and somewhat intriguing relationship between static semantics $\overline{T}$ and Clark's completions $comp(T)$ of finite belief theories. Note that we use here a slightly generalized version of Clark's completion introduced in (1; 4) (see Definition 3.2). Namely we first transform the theory T in a special normal form, called the residual theory. This transformed theory has exactlythe same models as T. Then we use the syntactic form of T to apply the original Clark's completion. Details will be given below.

It shows that the static semantics $\overline{T}$ of T is obtained by augmenting T with the set $\mathcal{B}comp(T) = \{\mathcal{B}F : F \in comp(T)\}$ thus ensuring that all formulae that belong to Clark's completion $comp(T)$ of T are believed to be true. It reduces reasoning under the static semantics $\overline{T}$ to the easily accomplished computation of Clark's completion $comp(T)$ together with theorem-proving in the underlying modal logic AEB, which can be done either by hand or using an automated theorem prover.

Both results open the way for more efficient implementation of static semantics: the first, because only one iteration is needed, and the second because reasoning in a non-standard logic (belief theories under static semantics) can be reduced to classical theorem proving involving Clark's completion.

2 Autoepistemic Belief Theories

We begin by recalling the definition and basic properties of *autoepistemic belief theories*, AEB, originally introduced in (22; 23). Consider a fixed propositional language $\mathcal{L}$ with standard connectives ($\vee$, $\wedge$, $\neg$, $\rightarrow$, $\leftarrow$, $\leftrightarrow$) and the propositional letters *true* and *false*. We denote the set of its propositions by $At_{\mathcal{L}}$. Extend the language $\mathcal{L}$ to a propositional modal language $\mathcal{L}_{AEB}$ by augment-

[3] To be precise, this requires T to be in a special normal form, called a residual program.

ing it with a modal operator $\mathcal{B}$, called the *belief* operator. The formulae of the form $\mathcal{B}F$, where F is an arbitrary formula of $\mathcal{L}_{AEB}$, are called *belief atoms* and are considered to be atomic formulae in the extended propositional modal language $\mathcal{L}_{AEB}$. The formulae of the original language $\mathcal{L}$ are called *objective.* Any propositional theory in the modal language $\mathcal{L}_{AEB}$ will be called an *autoepistemic belief theory* (or just a "belief theory" for short).

Definition 2.1 (Autoepistemic Belief Theories (23))
By a belief theory *we mean an arbitrary theory in the propositional language $\mathcal{L}_{AEB}$, i.e., a (possibly infinite) set of clauses of the form:*

$$B_1 \wedge \ldots \wedge B_m \wedge \mathcal{B}G_1 \wedge \ldots \wedge \mathcal{B}G_k \rightarrow A_1 \vee \ldots \vee A_l \vee \mathcal{B}F_1 \vee \ldots \vee \mathcal{B}F_n$$

where $m, n, k, l \geq 0$, A_i's and B_i's are objective atoms and F_i's and G_i's are arbitrary formulae of $\mathcal{L}_{AEB}$.

We assume the following two simple axiom schemata and one inference rule describing the arguably obvious properties of belief atoms:

(CA) Consistency Axiom: $\mathcal{B}(true)$ and $\neg\mathcal{B}(false)$
(DA) Distributive Axiom: $\mathcal{B}(F \wedge G) \leftrightarrow \mathcal{B}F \wedge \mathcal{B}G$
(IR) Invariance Inference Rule: $\frac{F \leftrightarrow G}{\mathcal{B}F \leftrightarrow \mathcal{B}G}$

Remark 2.1 Although the set of axioms used in the above definition is slightly different from the one used in the original definition of autoepistemic belief theories given in (23), the two axiom systems are completely equivalent (see Appendix, Section A.4.).

As usual, a mapping $\mathcal{I}$: $At_{\mathcal{L}} \cup \{\mathcal{B}F : F \in \mathcal{L}_{AEB}\} \rightarrow \{true, false\}$ denotes a propositional interpretation of $\mathcal{L}_{AEB}$, i.e. we simply treat the formulae $\mathcal{B}F$ as new propositions. Therefore, the notion of a model carries over from propositional logic. A formula $F \in \mathcal{L}_{AEB}$ is a propositional consequence of $T \subseteq \mathcal{L}_{AEB}$ iff for every interpretation $\mathcal{I}$: $\mathcal{I} \models T \implies \mathcal{I} \models F$. In the examples, we will represent models by sets of literals showing the truth values of only those objective and belief atoms which are relevant to our considerations.

Definition 2.2 (Derivable Formulae (23))
For any belief theory T we denote by $Cn_{AEB}(T)$ the smallest set of formulae of the language $\mathcal{L}_{AEB}$ which contains T, all (substitution instances of) the axioms (CA) and (DA) and is closed under both propositional consequence and the invariance rule (IR).

We say that a formula F is derivable *from the belief theory T if F belongs to $Cn_{AEB}(T)$. We denote this fact by $T \vdash_{AEB} F$. A belief theory T is* consistent *if $Cn_{AEB}(T)$ is consistent, i.e., if $Cn_{AEB}(T) \not\vdash_{AEB}$ false.*

From the results established in (23) it immediately follows that the following lemma, which will be needed in the sequel, holds.

Proposition 2.3 *For any belief theory T and formulae $F, G \in \mathcal{L}_{AEB}$:*

$$T \vdash_{AEB} (\mathcal{B}F \to \neg\mathcal{B}\neg F)$$
$$T \vdash_{AEB} (\mathcal{B}(F \to G) \wedge \mathcal{B}F) \to \mathcal{B}G$$
$$\text{If } T \vdash_{AEB} F \text{ then } T \vdash_{AEB} \mathcal{B}F.$$

The first statement says that if a formula F is believed to be true then its negation $\neg F$ is not. The second statement says that if the formulae $F \to G$ and F are believed then so is G. The last inference rule says that if a formula F is derivable then it is also believed.

2.1 Semantics of Autoepistemic Belief Theories

We now need to define a *non-monotonic semantics* for autoepistemic belief theories. Intuitively, we need to provide a meaning to the belief atoms $\mathcal{B}F$. We want the intended meaning of belief atoms $\mathcal{B}F$ to be based on the principle of *predicate minimization* (see (19; 16) and (18)): $\mathcal{B}F$ holds if F is minimally entailed, or, equivalently: $\mathcal{B}F$ holds if F is true in all minimal models. In order to make this intended meaning precise we first have to define what we mean by a minimal model of a belief theory.

Definition 2.4 (Minimal Models (23))
it contains the same belief atoms but has fewer objective atoms, i.e.

$$\{p \in At_{\mathcal{L}} \,:\, M \models p\} \subset \{p \in At_{\mathcal{L}} \,:\, N \models p\},$$
$$\{F \in \mathcal{L}_{AEB} \,:\, M \models \mathcal{B}F\} = \{F \in \mathcal{L}_{AEB} \,:\, N \models \mathcal{B}F\}.$$

By a minimal model *of a belief theory T we mean a model M of T with the property that there is* no *smaller model N of T. If a formula F is true in all minimal models of T then we write $T \models_{\min} F$ and say that F is* minimally entailed *by T.*

Example 2.5 Consider the following simple belief theory T:

$$\to Car.$$
$$Car \wedge \mathcal{B}\neg Broken \to Runs.$$

Let us prove that T minimally entails $\neg Broken$, i.e., $T \models_{min} \neg Broken$. Indeed, in order to find minimal models of T we need to assign an *arbitrary* truth value to the only belief atom $\mathcal{B}\neg Broken$, and then *minimize* the objective atoms $Broken$, Car and $Runs$. We easily see that T has the following two minimal models (truth values of the remaining belief atoms are irrelevant and are therefore omitted): $M_1 = \{\mathcal{B}\neg Broken,\ Car,\ Runs,\ \neg Broken\}$ and $M_2 = \{\neg\mathcal{B}\neg Broken,\ Car,\ \neg Runs,\ \neg Broken\}$. Since in both of them Car is true, and $Broken$ is false, we deduce that $T \models_{min} Car$ and $T \models_{min} \neg Broken$.

As in Moore's Autoepistemic Logic, the intended meaning of belief atoms in autoepistemic belief theories is enforced by defining suitable expansions of such theories.

Definition 2.6 (Static Expansions of Belief Theories (23; 22))
A belief theory $T^\diamond$ is called a static expansion *of a belief theory T if it satisfies the following fixed-point equation: $T^\diamond = Cn_{AEB}\Big(T \ \cup \{\mathcal{B}F \,:\, T^\diamond \models_{\min} F\}\Big)$, where F ranges over all formulae of $\mathcal{L}_{AEB}$.*

Thus a static expansion $T^\diamond$ of T is obtained by augmenting T with all belief atoms $\mathcal{B}F$ with the property that F is minimally entailed by the fixed point $T^\diamond$.

Theorem 2.7 (Least Static Expansion (23; 22))
Every belief theory T has the least *static expansion, namely, the least fixed point $\overline{T}$ of the operator $\Psi_T(X) Cn_{AEB}\Big(\ T \cup \{\mathcal{B}F : X \ \models_{\min} \ F\}\ \Big)$. Morever this belief closure operator Ψ_T is monotonic.*

More precisely, the least static expansion $\overline{T}$ of T can be constructed as follows. Let $T^0 = T$ and suppose that T^α has already been defined for any ordinal number $\alpha < \beta$. If $\beta = \alpha + 1$ is a successor ordinal then define:

$$T^{\alpha+1} = \Psi_T(T^\alpha) =_{def} Cn_{AEB}\Big(\ T \cup \{\mathcal{B}F : T^\alpha \ \models_{\min} \ F\}\ \Big),$$

where F ranges over all formulae in $\mathcal{L}_{AEB}$. Else, if β is a limit ordinal then define $T^\beta = \bigcup_{\alpha<\beta} T^\alpha$. The sequence $\{T^\alpha\}$ is monotonically increasing and has a unique fixed point $\overline{T} = T^\lambda = \Psi_T(T^\lambda)$, for some ordinal λ.

The above result allows us to establish the following useful characterization of the least static completion of a belief theory.

Theorem 2.8 (Characterization of Least Static Expansions)
The least static expansion of a belief theory T coincides with the smallest theory $\widehat{T}$ satisfying the conditions:

(1) $T \subseteq \widehat{T}$;
(2) $\widehat{T} = Cn_{AEB}(\widehat{T})$;

(3) if $\widehat{T} \models_{\min} F$ then $\mathcal{B}F \in \widehat{T}$.

Proof. Clearly the least static expansion $\overline{T}$ of T satisfies the above conditions 1–3 so $\widehat{T} \subseteq \overline{T}$. Moreover, from the conditions 1–3 and the very definition of Ψ_T it follows immediately that $\Psi_T(\widehat{T}) \subseteq \widehat{T}$. Since the operator Ψ_T is monotonic on such theories (see (23)), we conclude that $T^1 = \Psi_T(T) \subseteq \Psi_T(\widehat{T}) \subseteq \widehat{T}$, and, more generally, $T^\alpha \subseteq \widehat{T}$, for any ordinal α. We conclude therefore that $\overline{T} \subseteq \widehat{T}$. □

Observe that the *least* static expansion $\overline{T}$ of T contains those and only those formulae which are true in *all* static expansions of T. It constitutes the so called *static completion* of a belief theory T.

Example 2.9 Consider a slightly more complex belief theory T:

$$\begin{aligned} &\mathcal{B}\neg Broken \rightarrow Runs. \\ &\mathcal{B}\neg Fixed \quad \rightarrow Broken. \end{aligned}$$

In order to iteratively compute its static completion $\overline{T}$ we let $T^0 = T$. As in Example 2.5, one easily checks that $T^0 \models_{min} \neg Fixed$. Since $T^1 = \Psi_T(T^0) = Cn_{AEB}\Big(T \ \cup \ \{\mathcal{B}F : T^0 \models_{\min} F\}\Big)$, it follows that $\mathcal{B}\neg Fixed \in T^1$ and therefore $Broken \in T^1$. Since, $T^2 = \Psi_T(T^1) = Cn_{AEB}\Big(T \ \cup \ \{\mathcal{B}F : T^1 \models_{\min} F\}\Big)$, it follows that $\mathcal{B}Broken \in T^2$. From Proposition 2.3 we conclude that $\neg\mathcal{B}\neg Broken \in T^2$ and therefore $T^2 \models_{\min} \neg Runs$. Accordingly, since:

$$T^3 = \Psi_T(T^2) = Cn_{AEB}\Big(T \ \cup \ \{\mathcal{B}F : T^2 \models_{\min} F\}\Big),$$

we infer that $\mathcal{B}\neg Runs \in T^3$. As expected, the static completion $\overline{T}$ of T, which contains T^3, asserts that the car is considered not to be fixed and therefore broken and thus is not in a running condition.

Definition 2.10 (Static Completion and Static Semantics)
The least static expansion $\overline{T}$ of a belief theory T is called the static completion *of T. It describes the* static semantics *of a belief theory T.*

Consequently, like the predicate completion semantics of a logic program P is completely determined by its Clark's completion $comp(P)$, the static semantics of a belief theory T is fully determined by its static completion $\overline{T}$. It is easy to verify that a belief theory T either has a *consistent* static completion $\overline{T}$ or it does *not* have any consistent static expansions at all.

3 Static Completions vs. Clark's Completions

It is easy to see that if a belief theory is *finite* then the construction of its static completion (or the least static expansion) will stop after *countably* many steps. However, the surprising result obtained in this article shows that static completions of finite belief theories T are in fact obtained by means of a *single iteration* of the belief closure operator Ψ_T. This result eliminates the need for multiple iterations in the computation of static completions and allows us to replace the fixed-point definition of static completions by the equivalent explicit definition given by $\overline{T} = \Psi_T(T)$.

The second, closely related result establishes a very interesting and somewhat intriguing relationship between static completions $\overline{T}$ and Clark's completions $comp(T)$ of finite belief theories. It shows that the static completion $\overline{T}$ of a belief theory T coincides with $Cn_{AEB}\big(T \cup \{\mathcal{B}F : F \in comp(T)\}\big)$, i.e., with the set of formulae derivable from the belief theory T augmented with the set $\mathcal{B}comp(T) = \{\mathcal{B}F : F \in comp(T)\}$. The latter set represents the set of beliefs in formulae that belong to Clark's completion. Even though, strictly speaking, the last result only applies to the so called *residual theories* (see below), we do not lose generality, as we will show that every finite belief theory can be transformed into a finite residual belief theory with an equivalent set of minimal models. Since Clark's completion $comp(T)$ is easily computable, this result reduces reasoning under the static semantics to theorem-proving in the underlying modal logic AEB (this can be done either by hand or by using an automated theorem prover).

The proof of these two powerful results is based on the idea of adding to a belief theory T the set of formulae which ensure that models "seen" through the belief operator $\mathcal{B}$ are in fact minimal. As we will soon see, this task can be accomplished by a suitable generalization of Clark's completion but it only works for a restricted class of belief theories, namely those whose clauses do not have any objective premises. Such *residual theories* were previously introduced and investigated in the class of logic programs (6; 7; 11; 12; 4; 1). Here we give a slightly more general definition.

Definition 3.1 (Residual Belief Theories (1))
arbitrary belief theory whose clauses do not contain any objective (positive) premises, i.e., a (possibly infinite) set of arbitrary clauses $\mathcal{B}G_1 \wedge \ldots \wedge \mathcal{B}G_k \rightarrow A_1 \vee \ldots \vee A_l \vee \mathcal{B}F_1 \vee \ldots \vee \mathcal{B}F_n$*, where* $n, k, l \geq 0$*,* A_i*'s are objective atoms and* F_i*'s and* G_i*'s are arbitrary formulae of* $\mathcal{L}_{AEB}$*.*

Whenever convenient, clauses $\mathcal{B}G_1 \wedge \ldots \wedge \mathcal{B}G_k \rightarrow \mathcal{B}F_1 \vee \ldots \vee \mathcal{B}F_n$, i.e., clauses without any objective atoms in their heads, will be considered as clauses with a single objective atom $false$ in their head, i.e, clauses of the form: $\mathcal{B}G_1 \wedge$

$\ldots \wedge \mathcal{B}G_k \rightarrow false \vee \mathcal{B}F_1 \vee \ldots \vee \mathcal{B}F_n$.

Now we describe a natural extension of the notion of *Clark's completion* $comp(T)$. Clark's completion was initially introduced in (8) for the class of normal logic programs. Subsequently, its generalizations to disjunctive programs without positive premises were studied in (1). Here we extend it to the class of all residual belief theories. As usual, the goal is to ensure that an objective atom A is true only if it really has to be true, i.e., if there is a rule with A in the head, in which the body is true and all the other head literals are false.

Definition 3.2 (Extended Clark's Completion)
objective atom A from $\mathcal{L}$ (including the falsity atom $false$) we define the extended Clark's completion $comp(A, T)$ *of the atom A in T to be the formula:*

$$A \leftrightarrow \bigvee_{1 \leq m \leq s} (\mathcal{B}G_{1,m} \wedge \ldots \wedge \mathcal{B}G_{k_m,m} \wedge \neg A_{1,m} \wedge \ldots \wedge \neg A_{l_m,m} \\ \wedge \neg \mathcal{B}F_{1,m} \wedge \ldots \wedge \neg \mathcal{B}F_{n_m,m}),$$

where $\mathcal{B}G_{1,m} \wedge \ldots \wedge \mathcal{B}G_{k_m,m} \rightarrow A \vee A_{1,m} \vee \ldots \vee A_{l_m,m} \vee \mathcal{B}F_{1,m} \vee \ldots \vee \mathcal{B}F_{n_m,m}$, for $1 \leq m \leq s$, are all clauses in T containing the atom A in their heads.

By the Extended Clark's completion $comp(T)$ *of a belief theory T we mean the union of all completions $comp(A, T)$ of A in T, for all objective atoms A in the language $\mathcal{L}$ (including the falsity atom $false$).*

The following interesting result shows that if T is a residual belief theory, then the models of Clark's completion $comp(T)$ are precisely the minimal models of T itself. In other words, Clark's completion precisely describes the minimal model semantics of a residual belief theory T.

Proposition 3.3 *Let T be a finite residual belief theory. An interpretation M of the language $\mathcal{L}_{AEB}$ is a minimal model of T if and only if M is a model of Clark's completion $comp(T)$ of T.*

Proof. Contained in the Appendix B.1. □

It is worth pointing out that one of the weaknesses of the original version of Clark's completion proposed in (8) was the fact that it was applied not just to residual but rather to *arbitrary* normal programs. As a result, Clark's original completion did *not* enforce minimal model semantics, e.g., a tautology like $p \leftarrow p$ made the completion axiom for p useless. Another problem with the original definition of Clark's completion was that it did not distinguish between logical negation $\neg p$ and belief $\mathcal{B}\neg p$. As a result, Clark's completion could be inconsistent when rules like $p \leftarrow \mathcal{B}\neg p$ were present.

Although the above result, as well as the definition of Clark's completion, applies only to the class of residual belief theories, it turns out that *every* finite belief theory T can be transformed into a finite residual belief theory T_{res} so that their sets of minimal models coincide.

Proposition 3.4 *Every finite belief theory T can be transformed into a finite residual belief theory T_{res}, called the* residuum *of T, so that an interpretation M of the language $\mathcal{L}_{AEB}$ is a minimal model of T if and only if M is a minimal model of T_{res}.*

Proof. Contained in the Appendix B.2. □

In fact, just two elementary theory transformations, namely, "unfolding" (or partial evaluation, GPPE (4; 27)) and "elimination of tautologies", are sufficient to obtain such a residual program. For a detailed definition of these transformations the reader is referred to (1; 2; 4; 3). Here we describe them very briefly.

The "elimination of tautologies" allows us to remove a rule if it contains the same atom in both, head and body. Such rules are always trivially satisfied and therefore useless. The transformation of "unfolding" - also called $GPPE$ – is defined as follows. Suppose that B is an objective atom and let $B \vee \mathcal{A}_i \leftarrow \mathcal{R}_i$, $i = 1, \ldots, n$, be all rules about B, i.e., rules containing B in their head. The application of $GPPE$ to the rule $\mathcal{A} \leftarrow B \wedge \mathcal{R}$, that contains B in its body, results in a new belief theory obtained by deleting the rule $\mathcal{A} \leftarrow B \wedge \mathcal{R}$ and adding to the belief theory the new rules $\mathcal{A} \vee \mathcal{A}_i \leftarrow \mathcal{R} \wedge \mathcal{R}_i$.

We now give a simple example. Whenever convenient, we use the traditional inverse notation for clauses. Consider the following belief theory T. The first two clauses already have the required form. In the last clause, we unfold the body literal q and obtain the residual program T_{res}.

$$
\begin{array}{rlrl}
T: & p \vee q \leftarrow \mathcal{B}\neg r. & T_{res}: & p \vee q \leftarrow \mathcal{B}\neg r. \\
 & q \leftarrow \mathcal{B}\neg q. & & q \leftarrow \mathcal{B}\neg q. \\
 & r \leftarrow q. & & p \vee r \leftarrow \mathcal{B}\neg r. \\
 & & & r \leftarrow \mathcal{B}\neg q.
\end{array}
$$

One can easily see that this procedure represents a form of hyper-resolution.

As an immediate consequence of Propositions 3.3 and 3.4 we obtain the following important theorem which says that for any finite belief theory T, Clark's completion of its residuum T_{res} precisely describes the minimal model semantics of T.

Theorem 3.5 *Let T be any finite belief theory. An interpretation M of the language $\mathcal{L}_{AEB}$ is a minimal model of T if and only if M is a model of Clark's completion $comp(T_{res})$ of the residuum T_{res} of T.*

Finally, let us denote by $\mathcal{B}comp(T_{res})$ the set $\{\mathcal{B}F : F \in comp(T_{res})\}$. Intuitively, augmenting a given belief theory S with $\mathcal{B}comp(T_{res})$ ensures that all formulae that belong to Clark's completion $comp(T_{res})$ are believed. We will need the following lemma.

Lemma 3.6 *Let T be an arbitrary finite belief theory and let*

$$\widehat{T} = Cn_{AEB}\Big(T \cup \mathcal{B}comp(T_{res})\Big).$$

If a formula F is true in all minimal models of $\widehat{T}$ then $\mathcal{B}F$ is already contained in $\widehat{T}$. In other words, $\Psi_T(\widehat{T}) \subseteq \widehat{T}$.

Proof. Contained in the Appendix B.3. □

Now we are ready to state the first of the two fundamental results obtained in this article.

Theorem 3.7 (Static Completion vs. Clark's Completion)
The static completion $\overline{T}$ of a finite belief theory T can be equivalently defined as:

$$\overline{T} = Cn_{AEB}\Big(T \cup \mathcal{B}comp(T_{res})\Big).$$

Proof. Let $\widehat{T} = Cn_{AEB}\Big(T \cup \mathcal{B}comp(T_{res})\Big)$. We first show that $\widehat{T} = \Psi_T(T)$, i.e.

$$\widehat{T} = Cn_{AEB}\Big(T \cup \{\mathcal{B}F : T \models_{\min} F\}\Big).$$

- "$\subseteq$": Let $\mathcal{B}G \in \mathcal{B}comp(T_{res})$, i.e. $G \in comp(T_{res})$. By Theorem 3.5, $T \models_{\min} G$, and thus $\mathcal{B}(G)$ is contained in the right hand side.
- "$\supseteq$": Let $T \models_{\min} F$. By Theorem 3.5, $comp(T_{res}) \models F$. Then there are $G_1, \ldots, G_n \in comp(T_{res})$ such that $G_1 \rightarrow (G_2 \rightarrow (\ldots(G_n \rightarrow F)\ldots))$ is a propositional tautology. Thus $\mathcal{B}(G_1 \rightarrow (G_2 \rightarrow (\ldots(G_n \rightarrow F)\ldots)))$ is contained in $\widehat{T}$ (see Proposition 2.3). Since $\mathcal{B}(G_i) \in \widehat{T}$ by construction, we get $\mathcal{B}F \in \widehat{T}$ (again by Proposition 2.3).

From Lemma 3.6 it follows that $\Psi_T(\widehat{T}) \subseteq \widehat{T}$. Since $\widehat{T} = \Psi_T(T)$ we conclude that:

$$\Psi_T\Big(\Psi_T(T)\Big) \subseteq \Psi_T(T)$$

which means that $\Psi_T\Big(\Psi_T(T)\Big) = \Psi_T(T)$ (because Ψ_T is monotonically increasing by Theorem 2.7) and therefore $\Psi_T(\widehat{T}) = \widehat{T}$. This proves that $\widehat{T}$ is a fixed point of the belief closure operator Ψ_T and thus is a static expansion of T. Since $\widehat{T} = \Psi_T(T)$ we conclude that $\widehat{T}$ is the least such fixed point and thus coincides with the static completion $\overline{T}$ of T. □

The above result states that the static completion $\overline{T}$ of T is obtained by augmenting T with the set $\mathcal{B}comp(T_{res})$ thus ensuring that all formulae that belong to Clark's completion $comp(T_{res})$ of T_{res} are believed. It establishes an interesting and somewhat intriguing relationship between static completions and Clark's completions of finite belief theories. It also reduces reasoning under the static semantics (i.e., the computation of static completions $\overline{T}$) to the easily accomplished computation of Clark's completion $comp(T_{res})$ together with theorem-proving in the underlying modal logic.

As a consequence of the proof of the above result we immediately obtain the second of the two fundamental theorems obtained in this paper, namely a result stating that static completion $\overline{T}$ of an arbitrary finite belief theory T is always obtained by a *single iteration* of the belief closure operator $\Psi_T(T)$.

Theorem 3.8 (Explicit Characterization of Static Completions) *belief theory T is always obtained by a single iteration of the belief closure operator $\Psi_T(T)$: $\overline{T} = \Psi_T(T)$. In other words,*

$$\overline{T} = Cn_{AEB}\Big(T \cup \{\mathcal{B}F : T \models_{min} F\}\Big).$$

Proof. We have to compare the definition of the operator Ψ_T with the statement in Theorem 3.7. In particular we have to compare the set $\{\mathcal{B}F : T \models_{min} F\}$ with the set $\mathcal{B}comp(T_{res})$. By Proposition 3.3 they are identical. □

The above theorem states that the static completion $\overline{T}$ of T is obtained by augmenting T with beliefs $\mathcal{B}F$ in all formulae F which are true in all minimal models of the theory T itself. It eliminates therefore the need for multiple iterations in the computation of static completions and allows us to replace the fixed-point definition of static completions by the equivalent explicit definition. In addition to Theorems 2.7, 2.8 and 3.7, this is the fourth characterization of static completions presented in this paper. Needless to say, the existence of an equivalent *non-fixed point* definition of static completions significantly simplifies this notion and the underlying theory. It also provides the foundation for the results obtained in (5).

Example 3.9 Consider the belief theory T from Example 2.9. We already noted that $T \models_{min} \neg Fixed$ and thus

$$\mathcal{B}\neg Fixed \in \overline{T} = Cn_{AEB}\Big(T \cup \{\mathcal{B}F : T \models_{min} F\}\Big).$$

This implies that $\overline{T}$ contains also $Broken$. By Proposition 2.3, $\mathcal{B}Broken$ belongs to the static completion and so does $\neg\mathcal{B}\neg Broken$. From the same Proposition it follows that $\mathcal{B}\neg\mathcal{B}\neg Broken$ belongs to the completion.

Moreover, it is easy to verify that $T \models_{min} \mathcal{B}\neg Broken \leftrightarrow Runs$ (in fact, the formula $\mathcal{B}\neg Broken \leftrightarrow Runs$ belongs to Clark's completion of T) and therefore $T \models_{min} \neg\mathcal{B}\neg Broken \leftrightarrow \neg Runs$. It follows that $\overline{T}$ contains $\mathcal{B}(\neg\mathcal{B}\neg Broken \leftrightarrow \neg Runs)$, which can also be written in the form $\mathcal{B}((\neg\mathcal{B}\neg Broken \rightarrow \neg Runs) \wedge (\neg\mathcal{B}\neg Broken \leftarrow \neg Runs))$, and thus the distributive axiom DA can be applied:

$$\mathcal{B}(\neg\mathcal{B}\neg Broken \rightarrow \neg Runs) \wedge \mathcal{B}(\neg\mathcal{B}\neg Broken \leftarrow \neg Runs).$$

Now we can apply the second formula of Proposition 2.3 to the first conjunct (we already know that $\mathcal{B}\neg\mathcal{B}\neg Broken$ belongs to the completion). Therefore $\overline{T}$ also contains $\mathcal{B}\neg Runs$.

We again conclude that the static semantics of T derives $\mathcal{B}\neg Fixed$, $Broken$ and $\mathcal{B}\neg Runs$. Observe that while the above reasoning utilizes only one iteration of the belief closure operator, it is slightly more complex than the one used in Example 2.9 where several iterations were performed. Needless to say, both methods lead to identical results.

References

[1] Stefan Brass and Jürgen Dix. A General Approach to Bottom-Up Computation of Disjunctive Semantics. In J. Dix, L. Pereira, and T. Przymusinski, editors, *Nonmonotonic Extensions of Logic Programming*, LNAI 927, pages 127–155. Springer, Berlin, 1995.

[2] Stefan Brass and Jürgen Dix. Characterizations of the Disjunctive Stable Semantics by Partial Evaluation. *Journal of Logic Programming*, 32(3):207–228, 1997.

[3] Stefan Brass and Jürgen Dix. Characterizations of the Disjunctive Well-founded Semantics: Confluent Calculi and Iterated GCWA. *Journal of Automated Reasoning*, 20(1):143–165, 1998.

[4] Stefan Brass and Jürgen Dix. Semantics of (Disjunctive) Logic Programs Based on Partial Evaluation. *Journal of Logic Programming*, 38(3):167–213, 1999.

[5] Stefan Brass, Jürgen Dix, and Teodor. C. Przymusinski. Super Logic Programs. In L. C. Aiello, J. Doyle, and S. C. Shapiro, editors, *Principles of Kno]wledge Representation and Reasoning: Proceedings of the Fifth International Conference (KR '96)*, pages 529–541. San Francisco, CA, Morgan Kaufmann, 1996.

[6] François Bry. Logic programming as constructivism: A formalization and its application to databases. In *Proc. of the Eighth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems (PODS'89)*, pages 34–50, 1989.

[7] François Bry. Negation in logic programming: A formalization in constructive logic. In Dimitris Karagiannis, editor, *Information Systems and Artificial Intelligence: Integration Aspects*, pages 30–46, Berlin, 1990. Springer.

[8] Keith L. Clark. Negation as Failure. In H. Gallaire and J. Minker, editors, *Logic and Data-Bases*, pages 293–322. Plenum, New York, 1978.

[9] Y. Dimopoulos, B. Nebel, and J. Koehler. Encoding planning problems in non-monotonic logic programs. In *Proceedings of the Fourth European Conference on Planning*, Toulouse, France, September 1997. Springer-Verlag. To appear.

[10] Jürgen Dix, V.S. Subrahmanian, and George Pick. Meta-agent programs. *Journal of Logic Programming*, to appear, 1999.

[11] P. M. Dung and K. Kanchansut. A fixpoint approach to declarative semantics of logic programs. In E.L. Lusk and R.A. Overbeek, editors, *Proceedings of North American Conference Cleveland,Ohio, USA*. MIT Press, October 1989.

[12] P. M. Dung and K. Kanchansut. A natural semantics for logic programs with negation. In E.L. Lusk and R.A. Overbeek, editors, *Proceedings of the 9th Conference on Foundations of Software Technology and Theoretical Computer Science*, Berlin, 1989. Springer.

[13] Thomas Eiter, V. S. Subrahmanian, and Georg Pick. Heterogeneous Active Agents, I: Semantics. *Artificial Intelligence*, 108(1-2):179–255, 1999.

[14] T. Eiter and V. S. Subrahmanian. Heterogeneous Active Agents, II: Algorithms and Complexity. *Artificial Intelligence*, 108(1-2):257–307, 1999.

[15] Michael Gelfond and Vladimir Lifschitz. The Stable Model Semantics for Logic Programming. In R. Kowalski and K. Bowen, editors, *5th Conference on Logic Programming*, pages 1070–1080. MIT Press, 1988.

[16] Michael Gelfond, Halina Przymusinska, and Teodor Przymusinski. On the Relationship between Circumscription and Negation as Failure. *Artificial Intelligence*, 38:75–94, 1989.

[17] R.A. Kowalski. Predicate logic as a programming language. In *Proceeedings IFIP' 74*, pages 569–574. North Holland Publishing Company, 1974.

[18] John McCarthy. Circumscription: A Form of Nonmonotonic Reasoning.

Artificial Intelligence, 13:27–39, 1980.

[19] Jack Minker. On indefinite databases and the closed world assumption. In *Proceedings of the 6th Conference on Automated Deduction, New York*, pages 292–308, Berlin, 1982. Springer.

[20] R. C. Moore. Semantical Considerations on Nonmonotonic Logic. *Artificial Intelligence*, 25:75–94, 1985.

[21] Ilkka Niemelä and Patrik Simons. Efficient Implementation of the Well-founded and Stable Model Semantics. In M. Maher, editor, *Proceedings of the Joint International Conference and Symposium on Logic Programming*, pages 289–303, Bonn, Germany, September 1996. The MIT Press.

[22] Teodor C. Przymusinski. Static semantics for normal and disjunctive logic programs. *Annals of Mathematics and Artificial Intelligence*, 14:323–357, 1995. Special Issue on Disjunctive Programs.

[23] Teodor C. Przymusinski. Autoepistemic logic of knowledge and beliefs. *Artificial Intelligence*, 95:115–154, 1998.

[24] Prasad Rao, K. Sagonas, T. Swift, D. S. Warren, and J. Freire. XSB: A System for Efficiently Computing Well-Founded Semantics. In J. Dix, U. Furbach, and A. Nerode, editors, *Logic Programming and Non-Monotonic Reasoning, Proceedings of the Fourth International Conference*, LNAI 1265, pages 430–440, Berlin, June 1997. Springer.

[25] R. Reiter. On Closed-World Databases. In H. Gallaire and J. Minker, editors, *Logic and Data Bases*, pages 55–76. Plenum Press, New York, 1978.

[26] Raymond Reiter. Deductive question-answering on relational data bases. In Hervé Gallaire and Jack Minker, editors, *Logic and Data Bases*, pages 149–177, New York, 1978. Plenum.

[27] Chiaki Sakama and Hirohisa Seki. Partial Deduction of Disjunctive Logic Programs: A Declarative Approach. In *Logic Program Synthesis and Transformation – Meta Programming in Logic*, LNCS 883, pages 170–182, Berlin, 1994. Springer.

[28] Y. Shoham. Agent Oriented Programming. *Artificial Intelligence*, 60:51–92, 1993.

[29] V.S. Subrahmanian, Piero Bonatti, Jürgen Dix, Thomas Eiter, Sarit Kraus, Fatma Özcan, and Robert Ross. *Heterogenous Active Agents*. MIT-Press, 1999.

[30] Terry Swift. Using Tabling for nonmonotonic Programming. *Annals of Mathematics and Artificial Intelligence, Special Issue on* Logics in Artificial Intelligence*, edited by J. Dix and J. Lobo*, 1999.

[31] Allen Van Gelder, Kenneth A. Ross, and J. S. Schlipf. Unfounded Sets and well-founded Semantics for general logic Programs. In *Proceedings 7th Symposion on Principles of Database Systems*, pages 221–230, 1988.

A Proofs of Theorems from Section 3

A.1 Proof of Proposition 3.3

"$\Rightarrow$:" Suppose $\mathcal{I}$ is not a model of $comp(T)$, so there is a completion axiom $comp(A,T)$ of an atom A in T:

$$A \leftrightarrow \bigvee_{1\le m\le s} (\mathcal{B}G_{1,m} \wedge \ldots \wedge \mathcal{B}G_{k_m,m} \wedge \neg A_{1,m} \wedge \ldots \wedge \neg A_{l_m,m} \\ \wedge \neg \mathcal{B}F_{1,m} \wedge \ldots \wedge \neg \mathcal{B}F_{n_m,m}),$$

which is not satisfied in $\mathcal{I}$. If the direction $\leftarrow$ is not satisfied then $\mathcal{I}$ obviously is not even a model of T. So assume that this direction is satisfied for all completion axioms, i.e. $\mathcal{I}$ is a model of T. If the direction $\rightarrow$ is not satisfied then $\mathcal{I} \models A$ (consequently, $A \neq false$) but for every rule

$$\mathcal{B}G_{1,m} \wedge \ldots \wedge \mathcal{B}G_{k_m,m} \rightarrow A \vee A_{1,m} \vee \ldots \vee A_{l_m,m} \vee \mathcal{B}F_{1,m} \vee \ldots \vee \mathcal{B}F_{n_m,m},$$

with A in the head, one of the $A_{i,j}$ is true or one of the $\mathcal{B}F_{i,j}$ is true or one of the $\mathcal{B}G_{i,j}$ is false. But then the interpretation $\mathcal{I}'$, which differs from $\mathcal{I}$ only in $\mathcal{I}' \not\models A$, is also a model of T and therefore $\mathcal{I}$ is not minimal.

"$\Leftarrow$:" Suppose conversely that $\mathcal{I}$ is not a minimal model of T. If $\mathcal{I}$ is not even a model, i.e. does not satisfy some rule

$$\mathcal{B}G_{1,m} \wedge \ldots \wedge \mathcal{B}G_{k_m,m} \rightarrow A \vee A_{1,m} \vee \ldots \vee A_{l_m,m} \vee \mathcal{B}F_{1,m} \vee \ldots \vee \mathcal{B}F_{n_m,m},$$

then obviously the direction $\leftarrow$ of the completion axioms for every A_i is also violated. So let us assume that $\mathcal{I}$ is a model, but not minimal, so there is a smaller model $\mathcal{I}'$. We can choose $\mathcal{I}'$ in such a way that it is minimal. Let A be an objective atom with $\mathcal{I} \models A$ and $\mathcal{I}' \not\models A$ (consequently, $A \neq false$). Let us consider the completion axiom $comp(A,T)$ for A:

$$A \leftrightarrow \bigvee_{1\le m\le s} (\mathcal{B}G_{1,m} \wedge \ldots \wedge \mathcal{B}G_{k_m,m} \wedge \neg A_{1,m} \wedge \ldots \wedge \neg A_{l_m,m} \\ \wedge \neg \mathcal{B}F_{1,m} \wedge \ldots \wedge \neg \mathcal{B}F_{n_m,m}),$$

By the contraposition of the already proven opposite direction, the minimal model $\mathcal{I}'$ is a model of $comp(T)$, so it satisfies the completion axiom. Since $\mathcal{I}' \not\models A$, for every i there is a j such that $\mathcal{I}' \models A_{i,j}$ or $\mathcal{I}' \models \mathcal{B}F_{i,j}$ or $\mathcal{I}' \not\models \mathcal{B}G_{i,j}$. But since $\mathcal{I}'$ is smaller than $\mathcal{I}$, $\mathcal{I}' \models A_{i,j} \implies \mathcal{I} \models A_{i,j}$ and $\mathcal{I}' \models \mathcal{B}F_{i,j} \iff$

$\mathcal{I} \models \mathcal{B}F_{i,j}$ and $\mathcal{I}' \not\models \mathcal{B}G_{i,j} \iff \mathcal{I} \not\models \mathcal{B}G_{i,j}$. Thus, the right hand side is also false in $\mathcal{I}$, but A is true in $\mathcal{I}$, and therefore we conclude that $\mathcal{I} \not\models comp(T)$.□

A.2 Proof of Proposition 3.4

We will show that every belief theory T can be transformed into a residual belief theory T_{res} by applying just two transformations, namely, "elimination of tautologies" and "unfolding".

It is easy to see that each application of *GPPE*, perhaps combined with tautology elimination, removes a given objective atom B from the body of one clause without introducing any new clauses containing B in their bodies. This means that a finite number of such transformations will lead to a belief theory that does not contain B's in bodies of its clauses. Applying this procedure to all objective atoms appearing in a (finite) belief theory will therefore produce a residual belief theory.

Accordingly it suffices to establish the following lemma.

Lemma A.1 *Let T_1 and T_2 be two belief theories such that T_2 results from T_1 by an application of GPPE or deletion of tautologies. Then T_1 and T_2 have the same minimal models.*

The deletion of tautologies is an equivalence transformation, so it does not change the set of models, and thus also not the set of minimal models. Now suppose that we apply GPPE to a rule $\mathcal{A} \leftarrow B \wedge \mathcal{R}$, where B is an objective atom, and suppose that $B \vee \mathcal{A}_i \leftarrow \mathcal{R}_i$, $i = 1, \ldots, n$, are all the rules about B, i.e., rules containing B in their heads.

- Since the new rules $\mathcal{A} \vee \mathcal{A}_i \leftarrow \mathcal{R} \wedge \mathcal{R}_i$, resulting from the application of *GPPE*, are logical consequences of the rules in T_1, every model of T_1 is also a model of T_2.
- Let $\mathcal{I}$ be a minimal model of T_2. We first show that it is a model of T_1 (not necessarily minimal). Suppose this is not the case, so the rule $\mathcal{A} \leftarrow B \wedge \mathcal{R}$ is violated in $\mathcal{I}$ (this is the only rule in $T_1 \setminus T_2$). This means that $\mathcal{I} \models B$, $\mathcal{I} \models \mathcal{R}$, and $\mathcal{I} \not\models \mathcal{A}$. Since $\mathcal{I}$ is a minimal model of T_2, the interpretation $\mathcal{I}'$ with $\mathcal{I}' \not\models B$ (and otherwise equal to $\mathcal{I}$) cannot be a model of T_2. Since objective atoms do not occur negated in rule bodies, this means that there is a rule with B in the head, i.e. one of the rules $B \vee \mathcal{A}_i \leftarrow \mathcal{R}_i$ [4], such that $\mathcal{I}' \models \mathcal{R}_i$ and $\mathcal{I}' \not\models \mathcal{A}_i$. Since $\mathcal{I}' \models \mathcal{R}_i$, the atom B cannot be contained

[4] It is not a new rule $\mathcal{A} \vee \mathcal{A}_i \leftarrow \mathcal{R} \wedge \mathcal{R}_i$, because otherwise $\mathcal{A}$ would contain B, and the unfolded rule $\mathcal{A} \leftarrow B \wedge \mathcal{R}$ would be a tautology and could never be violated.

in $\mathcal{R}_i$, and since we exclude multiple occurrences in the head, it is also not contained in $\mathcal{A}_i$. Thus, $\mathcal{I} \models \mathcal{R}_i$ and $\mathcal{I} \not\models \mathcal{A}_i$. But now we infer that the new rule $\mathcal{A} \vee \mathcal{A}_i \leftarrow \mathcal{R} \wedge \mathcal{R}_i$ contained in T_2 is violated by $\mathcal{I}$, which is a contradiction.

We now have to show that the sets of minimal models are equal:

- Suppose that $\mathcal{I}$ is a minimal model of T_1. By the first part of this proof, we know that it is a model of T_2. If it were not minimal, there would have to be a smaller minimal model $\mathcal{I}'$ of T_2. But by the second part of this proof it would also have to be a model of T_1, contradicting the assumed minimality of $\mathcal{I}$ as a model of T_1.
- Suppose that $\mathcal{I}$ is a minimal model of T_2. Then it is a model of T_1. If there were a smaller model $\mathcal{I}'$ of T_1, it would also have to be a model of T_2, again leading to a contradiction. □

A.3 Proof of Lemma 3.6

First we prove the following lemma which characterizes minimal models of special belief theories.

Lemma A.2 *Let T be any belief theory and $T_{\mathcal{B}}$ be a set of formulae which contain only belief atoms. Minimal models of $T \cup T_{\mathcal{B}}$ are precisely those minimal models of T which satisfy $T_{\mathcal{B}}$.*

Proof.

- Let $\mathcal{I}$ be a minimal model of $T \cup T_{\mathcal{B}}$. Of course, $\mathcal{I}$ is a model of T and of $T_{\mathcal{B}}$. Now suppose that there is a smaller model $\mathcal{I}'$ of T. Since $\mathcal{I}$ and $\mathcal{I}'$ do not differ in the interpretation of belief atoms, $\mathcal{I}'$ is also a model of $T_{\mathcal{B}}$, and thus a model of $T \cup T_{\mathcal{B}}$. But this contradicts the assumed minimality of $\mathcal{I}$.
- Let $\mathcal{I}$ be a minimal model of T which also satisfies $T_{\mathcal{B}}$. Clearly, $\mathcal{I}$ is a model of $T \cup T_{\mathcal{B}}$. Since $\mathcal{I}'$ is also a model of T, the existence of a smaller model $\mathcal{I}'$ would contradict the minimality of $\mathcal{I}$. □

Now we continue the proof of of Lemma 3.6.

Lemma 3.6 *Let T be an arbitrary finite belief theory and let*

$$\widehat{T} = Cn_{AEB}\Big(T \cup \mathcal{B}comp(T_{res})\Big).$$

If a formula F is true in all minimal models of $\widehat{T}$ then $\mathcal{B}F$ is already contained *in $\hat{T}$. In other words, $\Psi_T(\hat{T}) \subseteq \hat{T}$.*

Let $T_{\mathcal{B}}$ be the subset of $\hat{T}$ which consists entirely of belief atoms, i.e. $\mathcal{B}\Big(comp(T_{res})\Big)$ together with the axioms (CA), all instances of the axiom (DA) and all formulae forced in by the derivation rule (IR). Then the models of $\hat{T}$ coincide with the models of $T \cup T_{\mathcal{B}}$ because the required closure under logical consequences does not change the set of models.

By Lemma A.2, minimal models of $T \cup T_{\mathcal{B}}$ are precisely those minimal models of T which satisfy $T_{\mathcal{B}}$. Suppose that F holds in all minimal models of $\hat{T}$. By Proposition 3.3, the minimal models of T coincide with the models of $comp(T_{res})$ and therefore F holds in all models of $comp(T_{res}) \cup T_{\mathcal{B}}$, i.e. it is a propositional consequence of $comp(T_{res}) \cup T_{\mathcal{B}}$. By the compactness of propositional logic, there is a finite subset $\{F_1, \ldots, F_n\} \subseteq comp(T_{res}) \cup T_{\mathcal{B}}$ such that $\{F_1, \ldots, F_n\} \models F$. But then $F_1 \wedge \cdots \wedge F_n \to F$ is a tautology and therefore, by Proposition 2.3, $\mathcal{B}(F_1) \wedge \cdots \wedge \mathcal{B}(F_n) \to \mathcal{B}(F)$ is contained in $\hat{T}$.

If F_i is from $comp(T_{res})$, then $\mathcal{B}(F_i)$ is obviously contained in $\hat{T}$ and otherwise we get $\mathcal{B}(F_i)$ by necessitation (see Proposition 2.3). Thus, by the closure under classical consequences, we conclude $\mathcal{B}(F) \in \hat{T}$. □

A.4 Proof of the equivalence of the axioms used in the definition of AEB to the original axioms

We show now that the axioms used in Definition 2.1 are completely equivalent to the following ones used in the original definition of autoepistemic belief theories given in (23):

(D) Consistency Axiom: $\neg\mathcal{B}false$
(K) Normality Axiom: $\mathcal{B}(F \to G) \;\to\; (\mathcal{B}F \to \mathcal{B}G)$
(N) Necessitation Inference Rule: $\frac{F}{\mathcal{B}F}$.

We recall that, in the presence of the axiom (K), the axiom (D) is equivalent to the axiom $\mathcal{B}F \to \neg\mathcal{B}F$, stating that if we believe in a formula F then we do *not* believe in F (see (23) for details).

Proposition A.3
The axioms (D) and (K) and the necessitation rule (N) are equivalent to the axioms (CA) and (DA) and the inference rule (IR).

More precisely, a theory in the language $\mathcal{L}_{AEB}$ contains all (substitution instances of) the axioms (D) and (K) and is closed under both standard propositional consequence and the necessitation rule (N) if and only if it contains *all (substitution instances of) the axioms (CA) and (DA) and is closed under both standard propositional consequence and the inference rule (IR).*

Proof. "$\Rightarrow$:" Suppose that a theory T contains all (substitution instances of) the axioms (D) and (K) and is closed under both standard propositional consequence and the necessitation rule (N). Since *true* belongs to T, from the necessitation rule (N) it follows that so does $\mathcal{B}true$ and thus the axiom (CA) holds.

To show that the axiom (DA) holds we first show that for any formulae F and G: $\mathcal{B}(F \wedge G) \rightarrow \mathcal{B}F \wedge \mathcal{B}G$ belongs to T. Clearly $(F \wedge G) \rightarrow F \in T$ and therefore, by the Necessitation Rule (N), $\mathcal{B}((F \wedge G) \rightarrow F) \in T$. From the Normality Axiom (K) we infer that $\mathcal{B}(F \wedge G) \rightarrow \mathcal{B}F \in T$. Similarly, $\mathcal{B}(F \wedge G) \rightarrow \mathcal{B}G \in T$. It follows that $\mathcal{B}(F \wedge G) \rightarrow \mathcal{B}F \wedge \mathcal{B}G \in T$.

To show that the axiom (DA) holds it now suffices to prove that for any formulae F and G: $\mathcal{B}F \wedge \mathcal{B}G \rightarrow \mathcal{B}(F \wedge G)$ belongs to T. Clearly, $F \rightarrow (G \rightarrow F \wedge G) \in T$ and therefore, by the Necessitation Rule (N), $\mathcal{B}(F \rightarrow (G \rightarrow F \wedge G)) \in T$. From the Normality Axiom (K) we infer that $\mathcal{B}F \rightarrow \mathcal{B}(G \rightarrow F \wedge G) \in T$. Applying the Normality Axiom (K) again we conclude that $\mathcal{B}F \rightarrow (\mathcal{B}G \rightarrow \mathcal{B}(F \wedge G)) \in T$. This shows that $\mathcal{B}F \wedge \mathcal{B}G \rightarrow \mathcal{B}(F \wedge G) \in T$.

To demonstrate that the inference rule (IR) holds suppose that $F \leftrightarrow G$ belongs to T. By necessitation, also $\mathcal{B}(F \leftrightarrow G)$ belongs to T. From the Normality axiom it follows immediately that so does $\mathcal{B}F \leftrightarrow \mathcal{B}G$.

"$\Leftarrow$:" Suppose that a theory T contains all (substitution instances of) the axioms (CA) and (DA) and is closed under both standard propositional consequence and the inference rule (IR). Clearly, the axiom (D) also holds. Since $(F \rightarrow G) \wedge F \leftrightarrow F \wedge G$ belongs to T, from the rule (IR) it follows that $\mathcal{B}((F \rightarrow G) \wedge F) \leftrightarrow \mathcal{B}(F \wedge G)$ belongs to T. From the axiom (DA) we conclude that also $\mathcal{B}(F \rightarrow G) \wedge \mathcal{B}(F) \leftrightarrow \mathcal{B}F \wedge \mathcal{B}G$ belongs to T. This proves that $\mathcal{B}((F \rightarrow G)) \wedge \mathcal{B}(F) \rightarrow \mathcal{B}(G)$ belongs to T and thus the Normality axiom (N) holds.

It remains to show that T is closed under necessitation. If F belongs to T then so does $F \leftrightarrow true$. By the rule (IR), the formula $\mathcal{B}F \leftrightarrow \mathcal{B}true$ belongs to T which implies that $\mathcal{B}F$ also belongs to T. $\Box$